A monaural masking release based on a similar mechanism as binaural unmasking

NICHOLAS R. CLARK, BARRIE A. EDMONDS, AND KATRIN KRUMBHOLZ

MRC Institute of Hearing Research, University Park, Nottingham, NG7 2RD, United Kingdom

Interaural temporal differences are known to play an important role in enhancing the detectability of signals in noise, but seem to have little contribution to the perceptual segregation of simultaneous sound sources. Conversely, pitch is commonly recognised for its role in simultaneous sound source segregation, while its potential effect on signal detectability has so far received little attention. The current study investigates the role of monaural pitch in signal detection with two experiments. The first experiment demonstrates that pitch cues can greatly enhance the detectability of signals in noisy situations, equalling the potential unmasking reported for binaural cues. This pitch-based unmasking effect is independent of the discriminability of the signal and masker pitches, suggesting that it is based on the reduction in the serial correlation of the composite stimulus induced by the signal. Similarly, in binaural unmasking, the detectability of the signal increases as a result of the decrease in interaural correlation that it produces. The second experiment, which was based on non-simultaneous masking, confirmed that the spectral resolution of the cochlear filters could only account for a small portion of the unmasking measured in the first experiment.

INTRODUCTION

The attributes of a tonal stimulus that give rise to the perception of pitch are thought to be processed using a combination of two mechanisms. A temporal mechanism, which evaluates the serial correlation of the stimulus, and a spectral mechanism, which evaluates the peaks in the distribution of time-averaged activity across the tonotopic array (for review, see de Cheveigné, 2005). Temporal pitch models and models of binaural temporal processing (for review, see Colburn, 1996) are similar, in that they are both based on a form of correlation analysis. Therefore, pitch cues may be expected to give rise to similar unmasking effects as binaural temporal cues. Nevertheless, pitch cues have received little attention with regard to their role in signal detection. This is, because summation of a periodic signal and masker can give rise to periodic temporal envelope interactions, which may be perceived as beating or roughness. Therefore, the separate contributions from pitch-related and envelope-related cues to any observed masking release would seem difficult to disentangle (Micheyl *et al.*, 2006). The aim of the current study was to test whether pitch cues provide a similar level of unmasking as binaural cues, by using a tonal stimulus with a stochastic envelope waveform.

Proceedings of ISAAR 2009: Binaural Processing and Spatial Hearing. 2nd International Symposium on Auditory and Audiological Research. August 2009, Elsinore, Denmark. Edited by J. M. Buchholz, T. Dau, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN: 87-990013-2-2. The Danavox Jubilee Foundation, 2010.

DETECTION EXPERIMENT

Methods

The stimuli were iterated ripple noises (IRNs), generated using 16 iterations of the add-original algorithm (Yost *et al*., 1996). The algorithm involves delaying a copy of a noise by a delay, *d*, adding the copy back to the original and iterating the process. IRNs have a serial correlation with period *d*, and their spectra exhibit harmonic peaks at integer multiples of 1/*d*, henceforth referred to as the repetition rate. The IRN stimuli were filtered into one of two spectral regions. The passband of the low region ranged from 0.78 to 2.98 kHz, and from 2.64 to 4.84 kHz in the high region. If pitch processing is based on a similar mechanism as the binaural masking release, then a pitch-related release from masking would occur irrespective of harmonic resolvability. Therefore, repetition rates of the IRN signals were chosen according to the rule of Shackleton and Carlyon (1994), to include both unresolved (53.03 Hz and 150.00 Hz) and resolved (150.00 Hz and 424.26 Hz) IRNs in both spectral bands. Thresholds were measured for an IRN signal in the presence of an IRN masker as a function of the difference in repetition rate between the signal and masker IRNs. The repetition rate differences were quantified using a logarithmic unit of measure, cents, where one cent is equal to the ratio $2^{1/1200}$. Thresholds were measured for masker repetition rates that were 0.00, -5.00, -12.57, -31.62, -79.53 and -200.00 cents relative to each signal repetition rate. The stimuli were generated digitally with a 25-kHz sampling rate and a 24-bit amplitude resolution using Matlab and TDT System 3. Detection thresholds were measured with an adaptive 3I-3AFC procedure using a 2-down 1-up rule, where signal level was the adaptive parameter. The stimuli were presented in a continuous equally-exciting noise to mask audible distortion products. The noise was lowpass filtered at 0.5 octaves below the lower cutoff frequency of the stimulus passband using an 8^{th} order Butterworth filter and had a level of about 50 dB SPL per equivalent rectangular bandwidth (ERB) (Glasberg and Moore, 1990).

Results

The data in the left-hand panel of Fig. 1, show that the difference between the repetition rates of the signal and masker IRNs had a strong influence on the signal detection threshold. When the signal and masker IRNs had equal repetition rate, the loudness difference between the masker alone and signal plus masker intervals was the only detection cue. When a rate difference was introduced between signal and masker, the loudness cue remained, but there were also an additional pitch-related cue: At small rate differences, listeners reported using a reduction in overall pitch strength in the signal interval. At larger pitch differences, listeners were able to perceptually segregate the signal and masker components based on their pitches and thus 'hear out' the signal component. These additional cues improved detection performance by up to 15 dB for the resolved stimuli. Importantly, a masking release of almost the same magnitude was observed for the unresolved stimuli, for which spectral cues were supposedly unavailable. This suggests that the masking release was based on temporal cues, probably involving a mechanism analogous to the correlational mechanism thought to underlie binaural unmasking.

Modelling

In order to test whether the masked thresholds can be accounted for by temporal interactions alone, we simulated the data using a standard temporal pitch model. The model was based on the auditory image model (Patterson *et al*., 1992), which was implemented in three stages. The first stage was a 49-channel gammatone filterbank (gtf) with center frequencies between 0.2 and 6 kHz, evenly distributed on the ERB scale at approximately 2 channels per ERB. The second stage used half-wave rectification, log compression and a 2nd-order lowpass filter with a 1.2-kHz cutoff frequency to simulate the mechanical-to-neural transduction performed by the inner hair cells. Finally, a channel-by-channel time-interval histogram of the neural activity in each channel was produced via strobed temporal integration (sti). The histograms were then summed across channels and normalised by the peak at zero lag to remove level information to produce a summary auditory image (SAI). The Euclidean distance, *D*, was used to measure the differences between the SAIs for the signal plus masker and masker alone. *D* is the square-root of the integral of the squared differences between the SAIs, so it includes differences at all time intervals within the summary images. For each condition, *D* was calculated as a function of the signal-to-masker ratio (SMR). The simulated threshold was defined as the SMR at which *D* reached a criterion level, *C*, and this criterion was the main parameter in the fitting process. All conditions were fitted simultaneously, with a fixed value of *C*, which was then varied to find the value that minimized the root-mean-square (RMS) error between the simulated and observed thresholds. The simulation results are shown in the central panel of Fig. 1. While the simulations captured the general pattern of the listener data remarkably well, we were curious as to whether a more sophisticated model of the auditory periphery would be able to further increase the accuracy of the fit. For that, we replaced the gtf and instantaneous logarithmic compression used in the original version of the model with the pole-zero filter cascade (pzfc) filterbank, which implements compression dynamically (Walters *et al*., 2009). The pzfc filterbank had the effect of reducing the background level in the time interval histograms, and thereby afforded a considerable increase in the accuracy of the model predictions (right panel in Fig. 1). The use of the pzfc allowed even the detailed features of the threshold pattern, such as the difference between the two unresolved conditions and the difference between the asymptotic thresholds of resolved and unresolved data, to be captured.

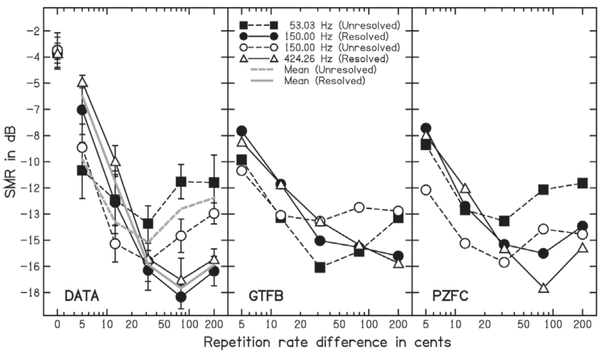

Fig. 1: The left panel shows listener data averaged across 5 participants. Error bars represent standard error. The parameter is the signal repetition rate. Simulated thresholds are shown when the gammatone filterbank (gtfb) was used (centre), and when the pole-zero filter cascade (pzfc) filterbank was used (right).

PULSATION THRESHOLD EXPERIMENT

In the first experiment, a large masking release was observed even for the unresolved stimuli, for which spectral cues were assumed to be unavailable based on Shackleton and Carlyon's (1994) definition of the limit of harmonic resolvability. In order to gain a more quantitative estimate of the contribution of spectral cues to our detection results, we used the pulsation threshold technique invented by Houtgast (1972) to measure non-simultaneous masked thresholds for IRN signals in IRN maskers. We expected spectral contributions to be modest, as the model described in the previous experiment accurately predicted simultaneous masked thresholds purely based on the temporal interactions between the signal and masker IRNs. These temporal interactions were prevented in this experiment by presenting signal and masker IRNs non-simultaneously.

Methods

Signal and masker IRNs were generated and filtered in the same way as in the previous experiment. Stimuli were composed of sequences of temporally interleaved signal and masker IRNs (Fig. 2) with overlapping 5-ms squared-cosine cross-faded ramps to prevent audible clicks at the transitions. Pulsation thresholds were measured using the *doublet* procedure (Bode and Carhart, 1973; Leek, 2001), where each run consisted of 2 interleaved adaptive 2I-2AFC tracks. As before, the adaptive parameter was the level of the signal IRN. The final threshold estimates from each track were averaged to target the 50% point on the psychometric function of the continuity judgment.

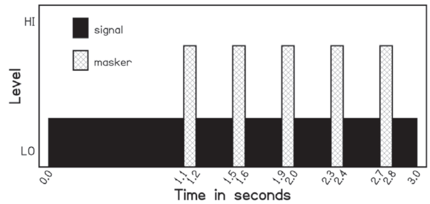

Fig. 2: Schematic of the sequence structure of the pulsation stimuli.

Results

Pulsation thresholds for the unresolved stimuli remained constant, irrespective of the repetition rate difference between the signal and masker IRNs (Fig. 3), as would be expected if the unresolved stimuli had a flat internal spectrum. For the resolved stimuli, the lowest pulsation thresholds occurred when there was a 79-cents repetition-rate difference between the signal and masker IRNs. At this point, the unmasking contribution from the spectral separation of signal and masker IRNs was the greatest. The difference between the pulsation thresholds for the resolved and unresolved stimuli at the 79-cents repetition-rate difference suggests that the spectral contribution to unmasking was at most 4 dB, which is much less than the 15-dB masking release observed in the simultaneous masking experiment. This indicates that the majority of the masking release observed in this study is based on temporal pitch cues.

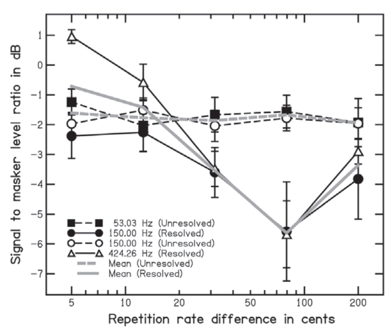

Fig. 3: Pulsation thresholds averaged across 5 participants. Error bars represent standard error.

CONCLUSIONS

The current results indicate that, like binaural temporal cues, pitch can markedly increase the detectability of signals in noisy environments, producing large masking level differences (MLDs) for both resolved and unresolved stimuli. Modelling showed this masking release can be accounted for by the reduction in serial correlation induced by the signal. A pulsation threshold experiment confirmed that, even for the resolved stimuli, spectral differences between the signal and masker contributed little to the pitch-related masking release, indicating that the masking release was mainly based on temporal information. Taken together, these results suggest that pitch-related and binaural masking release are based on a similar temporal correlation mechanism.

REFERENCES

Bode, D. L., and Carhart, R. (**1973**). "Measurement of articulation functions using adaptive test procedures," IEEE Transactions on Audio and Electroacoustics, **AU21**, 196-201.

Colburn, H. S. (**1996**). "Computational Models of Binaural Processing," in *Auditory Computation*, edited by H. L. Hawkins, T. A. McMullen, A. N. Popper, and R. F. Fay (Springer, New York), pp. 332-400.

de Cheveigné, A. (**2005**). "Pitch perception models," in *Pitch: Neural Coding and Perception*, edited by C. J. Plack, and A. J. Oxenham (Springer Verlag, New York).

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of Auditory Filter Shapes from Notched-Noise Data," Hear. Res., **47**, 103-138.

Houtgast, T. (**1972**). "Psychophysical evidence for lateral inhibition in hearing," J. Acoust. Soc. Am., **51**, 1885-1894.

Leek, M. R. (**2001**). "Adaptive procedures in psychophysical research," Perception and Psychophysics **63**, 1279-1292.

Micheyl, C., Bernstein, J. G. W., and Oxenham, A. J. (**2006**). "Detection and F0 discrimination of harmonic complex tones in the presence of competing tones or noise," J. Acoust. Soc. Am., **120**, 1493-1505.

Patterson, R., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. H. (**1992**). "Complex sounds and auditory images," in *Auditory physiology and perception*, edited by Y. Cazals, L. Demany, and K. Horner (Oxford), pp. 429-446.

Shackleton, T. M., and Carlyon, R. P. (**1994**). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," J. Acoust. Soc. Am., **95**, 3529-3540.

Walters, T. C., Patterson, R., and Lyon, R. F. (**2009**). "The Pole-Zero Filter Cascade, a dynamic compressive auditory filterbank," in *European career workshop for PhD students in hearing research*, Nottingham, UK.

Yost, W. A., Patterson, R., and Sheft, S. (**1996**). "A time domain description for the pitch strength of iterated rippled noise," J. Acoust. Soc. Am. **99**, 1066-1078.